Improving Taxonomic Name Finding in the Biodiversity Heritage Library

Joel Richard ‡

‡ Smithsonian Libraries, Washington, United States of America

Abstract

As the world's largest open access digital library for biodiversity literature and archives, the Biodiversity Heritage Library (BHL) provides access to over a quarter-million volumes of natural history literature to researchers around the world. One of its services is to index taxonomic names in the collection to allow researchers to locate publications about specific taxa. The Global Names Architecture (GNA) is a system of web services to register, find, index, check and organize biological scientific names. GNA recently developed a new Name Finding algorithm and tool that has been integrated with BHL to improve taxonomic name searches within BHL. In our presentation, we will discuss a brief history of name finding in BHL, development of the Name Finding algorithm, results from implementing the algorithm, and challenges that still await us in the realm of taxonomic name finding in BHL.

Keywords

taxonomic name identification, OCR, optical character recognition

Presenting author

Joel Richard

Presented at

TDWG 2020